Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2484508 A1 2003/12/04

(21) 2 484 508

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2003/05/16
(87) Date publication PCT/PCT Publication Date: 2003/12/04
(85) Entrée phase nationale/National Entry: 2004/10/28
(86) N° demande PCT/PCT Application No.: GB 2003/002125
(87) N° publication PCT/PCT Publication No.: 2003/101051
(30) Priorité/Priority: 2002/05/23 (10/154,308) US

(51) Cl.Int.^{7}/Int.Cl.^{7} H04L 12/56, H04L 12/43

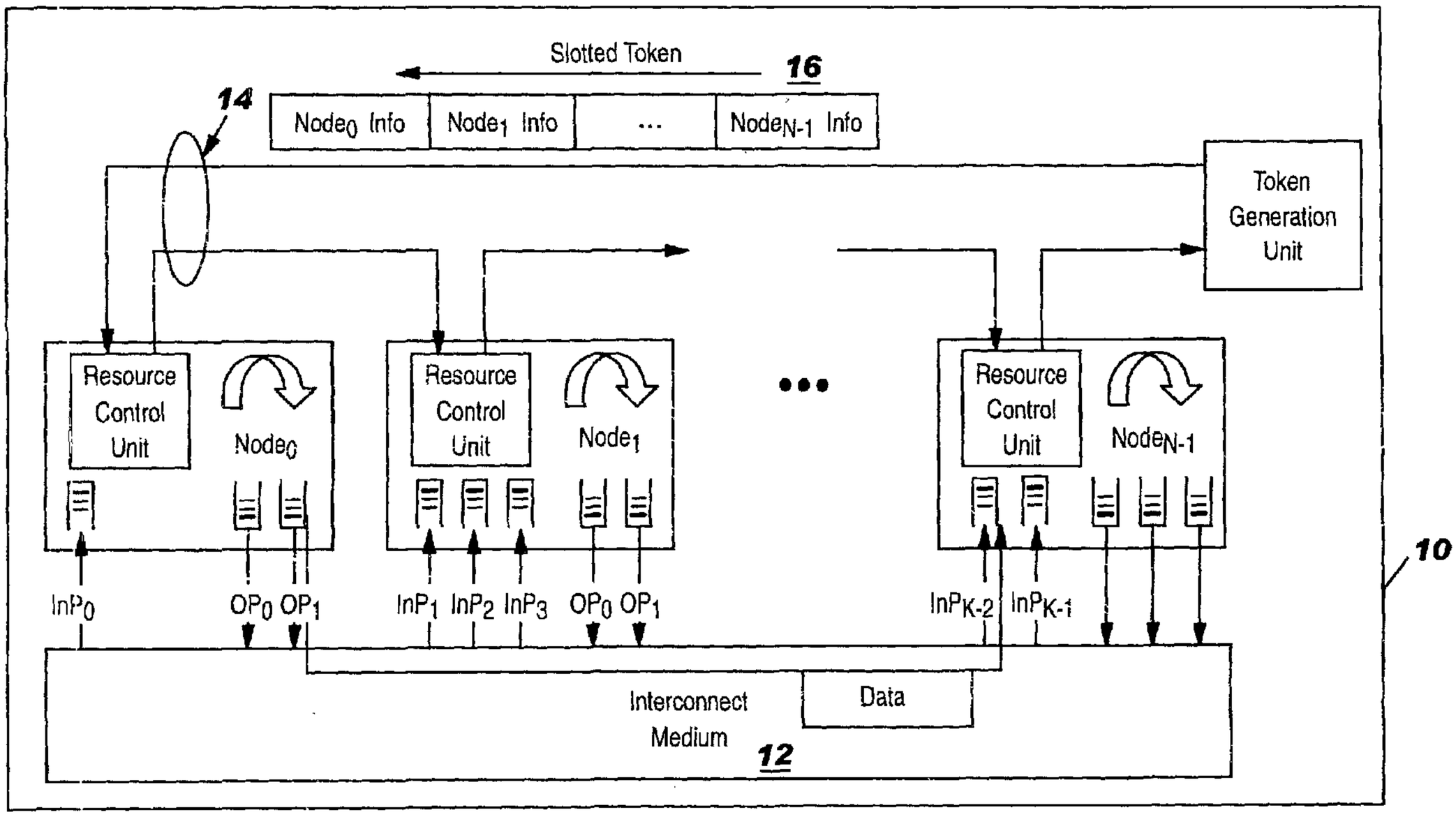
(71) Demandeur/Applicant:
INTERNATIONAL BUSINESS MACHINES CORPORATION, US

(72) Inventeurs/Inventors:
PEYRAVIAN, MOHAMMAD, US;
RINALDI, MARK ANTHONY, US;
SIEGEL, MICHAEL STEVEN, US;
SABHIKHI, RAVINDER KUMAR, US

(74) Agent: ROSEN, ARNOLD

(54) Titre : APPAREIL, PROCEDE ET PROGRAMME INFORMATIQUE POUR RESERVER DES RESSOURCES DANS DES SYSTEMES DE COMMUNICATION

(54) Title: APPARATUS, METHOD AND COMPUTER PROGRAM TO RESERVE RESOURCES IN COMMUNICATIONS SYSTEM



(57) Abrégé/Abstract:
A Resource Reservation System includes a Token Generation Unit (TGU) which generates and circulates among nodes of a communications system a Slotted Token (SLT) message having sub-fields to carry identification number for each input port in a node and the resource available for each input port. On receiving the message the Resource Control Unit (RCU) in each node can write port identification number, available resource in appropriate sub-fields of the SLT message, and reserve resources in other nodes by adjusting information in the sub-field associated with the other nodes.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 December 2003 (04.12.2003)

PCT

(10) International Publication Number
WO 03/101051 A1

(51) International Patent Classification[7]: H04L 12/56, 12/43

(21) International Application Number: PCT/GB03/02125

(22) International Filing Date: 16 May 2003 (16.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/154,308 23 May 2002 (23.05.2002) US

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, NY 10504 (US).

(71) Applicant *(for MG only)*: **IBM UNITED KINGDOM LIMITED** [GB/GB]; PO Box 41, North Harbour, Portsmouth, Hampshire PO6 3AU (GB).

(72) Inventors: **PEYRAVIAN, Mohammad**; 304 Oxcroft St., Morrisville, NC 27560 (US). **RINALDI, Mark, Anthony**; 1201 Queensbury Circle, Durham, NC 27713 (US). **SIEGEL, Michael, Steven**; 10625 Lowery Drive, Raleigh, NC 27615 (US). **SABHIKHI, Ravinder, Kumar**; 232 Strathburgh Lane, Cary, NC 27511 (US).

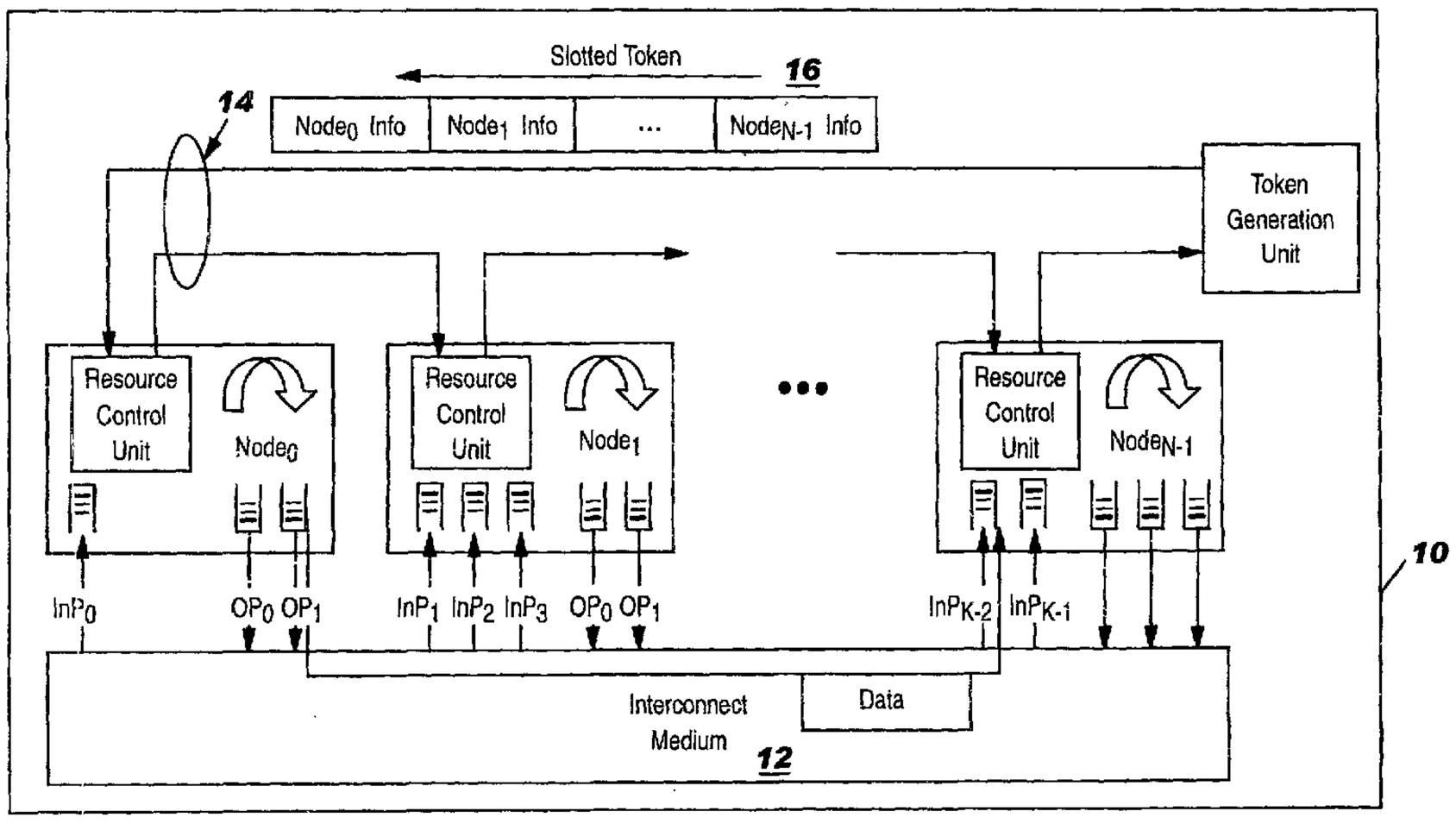
(74) Agent: **BURT, Roger, James**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester, Hampshire SO21 2JN (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: APPARATUS, METHOD AND COMPUTER PROGRAM TO RESERVE RESOURCES IN COMMUNICATIONS SYSTEM



(57) Abstract: A Resource Reservation System includes a Token Generation Unit (TGU) which generates and circulates among nodes of a communications system a Slotted Token (SLT) message having sub-fields to carry identification number for each input port in a node and the resource available for each input port. On receiving the message the Resource Control Unit (RCU) in each node can write port identification number, available resource in appropriate sub-fields of the SLT message, and reserve resources in other nodes by adjusting information in the sub-field associated with the other nodes.

WO 03/101051 A1

WO 03/101051 A1

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

APPARATUS, METHOD AND COMPUTER PROGRAM TO RESERVE RESOURCES IN COMMUNICATIONS SYSTEM

The present invention relates to communications systems in general and in particular to resource reservations within said communications systems.

A conventional communications system or network is comprised of a plurality of nodes coupled by an interconnect medium. Communication is effected by one node termed "Source Node" sending data to another node termed the "Destination Node". In order to maintain a particular Quality of Service (QoS) the Destination Node must reserve sufficient resources to process the data without undue delay. In fact, not only must the destination node reserve sufficient resources but any intermediate node that the data must traverse before reaching its destination must also reserve sufficient resources to ensure prompt processing of the data within these intermediate nodes. For example, the nodes must have enough storage space to buffer the data before processing. If adequate buffering is not available the node may have to discard the data. In this example storage is a resource. But, in general, a resource can be anything required to receive and process data. As a consequence a resource may include memory space, processor cycle, link, bandwidth etc.

The prior art provides several flow control proposals for managing data flow within communications networks. Resource management is an integral part of the flow control method. The prior art flow control proposals include Braden et al; Resource Reservation Protocol (RSVP), IETF RFC 2205, September 1997. The RSVP provides for receiver- initiated setup of resource reservation. In other words the destination node reserves resources based on a message sent by a source node. The RSVP protocol can be used by a host to request bandwidth from the network for data flows. The RSVP is usually used by routers to deliver bandwidth requests to all nodes along the path or paths of a flow. The node issuing the RSVP can also request confirmation assuring that the request has been installed in the network. One of the drawbacks is that the RSVP protocol reserves resources only for simpler flows. Stated another way, RSVP requests resources in only one direction. To this end RSVP treats a sender (source) as logically distinct from a receiver (destination) even though the same application process may act simultaneously as both a sender and receiver.

In another flow control scheme termed "Rate Base" the rate at which data is permitted to be delivered from a source to a destination is controlled via a feedback signal from destination to source. If resources are available at the destination the source may transmit data without restriction. If resources at the destination are in short supply or unavailable the rate of transmission is restricted to the point of cut off altogether. One such rate base technique is described in an ATM document #94-0735 entitled "Enhanced Proportional Rate Control Algorithm" by Larry Roberts, August 1994.

In yet another flow control scheme termed "Credit Base Control", a destination node generates and forwards "credits" to the Source node which may only transfer data if it has credits outstanding. The credits reflect the ability of the destination node to handle data. One such credit base controlled system is described in an ATM Forum document #94-0632 entitled "Credit-Based Proposal for ATM Traffic Management by Hunt et al., July 1994.

The prior art discusses the relationship between two nodes and the allocation of resources by a destination node for use by a source node, but provides no way to extend this to reserving resources throughout a plurality of interconnected nodes.

The invention provides a method as claimed in claim 1.

Preferably, the Resource Reservation System of the present invention includes a Token Generation Unit which generates a special message termed a "Slotted Token (SLT)" which is transmitted to all nodes in the network. The SLT includes a plurality of sub-fields with each sub-field relating to a node in the network. Each sub-field carries an identification (ID) for each input port at the node and a value indicating resources available at the port.

Preferably, each node is provided with a Resource Control Unit (RCU) that monitors input ports in the node and communicates via the SLT, to other nodes the available resources for each of the input ports. The RCU also reserves resources in other nodes to which said RCU may wish to send data. The SLT may be circulated in a path dedicated to transmit the SLT or in the interconnecting path which transmits data between the nodes. On its first pass the RCU in each node enters (write) the available resources for the input port in the space reserved in the SLT for that input port. Upon writing the information for all of its input ports, the SLT is

forwarded to another node which does the same. The process continues until all nodes in the network make entries in the SLT. On the first or subsequent pass of the SLT each RCU reserves the resource it needs in a particular input port by subtracting the resource from the value recorded in the space associated with the particular port.

The present invention therefore provides coordination between a plurality of nodes because a message is used to transmit information regarding resource availability between all nodes, and nodes are able to reserve resources by adjusting the information regarding resources availability in the message.

Preferred embodiments of the present invention will now be described in detail by way of example only with reference to the following drawings:

Figure 1 shows a block diagram of the Decentralized Out-of-Band Resource Reservation system according to the teachings of the present invention;

Figure 2 shows a graphical representation of the Slotted Token (SLT) format according to the teachings of the present invention;

Figure 3 shows a flowchart for logic in the Token Generation Unit (TGU);

Figure 4 is a flowchart for logic in the Resource Control Unit (RCU); and

Figure 5 shows a block diagram of the Decentralized In-band Resource Reservation Unit according to the teachings of the present invention.

To simplify the description common elements are identified by the same name, numeral or other symbols in the figures.

Figure 1 shows a block diagram of a Resource Reservation Communications System according to teachings of the present invention. The Resource Reservation Communications System includes a communications subsystem and a resource reservation subsystem. The communications subsystem includes Node:0, Node:1 . . . Node:(N-1). The nodes are coupled together by interconnect medium 12. The communications subsystem may take a plurality of different forms. For example, the communications subsystem can be a box such as a router with each node being a blade in the router.

In such an embodiment the interconnecting medium 12 could be a backplane in the router carrying a bus or optical channel for transmitting data between the respective blades. Likewise, the communications subsystem could be a plurality of boxes, each box representing a node, interconnected by an interconnecting medium 12 such as a local are network (LAN) or other types of communications highway such as Internet, etc. Stated another way, the communications subsystem can be any network in which data has to be transferred from one unit to another unit in the network.

Referring to Figures 1 and 5, each node in the communications subsystem has one or more input ports and one or more output ports. In particular, Node:0 has an input port labelled InP:0 and output ports labelled OP:0 and OP:1. In a similar manner Node:1 and Node:(N-1) have appropriate input and output ports labelled as shown in the figure. As a general principle the direction of data flow in the resource reservation communications system 10 is shown by the arrows. Data transmission between nodes such as Node:0 and Node:(N-1) is transported along the interconnecting medium 12. Each input port and output port is provided with a buffer shown as a 3-sided symbol in the figures. The horizontal lines in the 3-sided symbol represent a stack or queue of data which is placed in the buffer. Other types of symbols can also be used to represent the buffering.

Still referring to Figures 1 and 5, the resource reservation subsystem includes a resource control unit embedded in each of the nodes and a token generation unit interconnected by communications media 14 (Figure 1) or interconnect medium 12 (Figure 5). The communications media 14 can be any transmission medium on which a message termed "Slotted Token" (to be discussed hereinafter) generated by the token generation unit is transmitted. It should be noted that in Figure 1 the Slotted Token is transmitted on a dedicated transmission path such as communications media 14 whereas in Figure 5 the Slotted Token is transmitted on the interconnect medium 12 which also transmits the data. The functions which are performed by the Resource Control Unit (RCU) includes monitoring the input ports in the node in which the RCU is embedded and communicates to other nodes the available resources for its input ports. The RCU is also in charge of reserving resources in other nodes to which it needs to send data for processing. The RCU reserves resources in another node by subtracting the desired amount from the value carried in the Slotted Token (SLT) for the particular node. The token generation unit (TGU) generates a special message termed "Slotted Token"

which is transmitted in turn to all the resource control units in the system. Even though the token generation unit is shown as a separate unit in Figures 1 and 5 the function of the TGU can be integrated with the nodes RCU thereby eliminating the need for a separate TGU.

Referring now to Figures 1, 2 and 5, the Slotted Token 16 includes a plurality of sub-fields each of which carries information relative to a node in the system. Turning to Figure 2 for the moment, the first sub-field labelled Node:0 Info carries information relative to Node:0. Likewise, the sub-field labelled Node:1 Info carries information for Node:1 and so forth. The information which is in the sub-field includes indicia representing the identification (i.d.) of the input ports associated with that node and the resources available at that input port. With particular reference to Figure 2 the first sub-field for Node:0 has partition labelled InP:0 which carries the identification of that port and the partition labelled $AvResInP_0$ carries the resource available at that input port. Likewise, for $Node_1$ there are three input ports labelled InP:1 InP:2 InP:3. With the available resources for each of the input ports recorded in space adjacent to the input port ID. With this Slotted Token message being circulated a resource control unit can enter the port number and associated resource in the space allotted for that port. Likewise, the resource control unit can reserve resources in other ports by adjusting the available resource to indicate the resource that the node wants another node to reserve in order to process data from the requesting node. The partitioning of Node:(N-1) is similar to the other nodes and will not be described further.

Figure 3 shows a flowchart illustrating the operation of the token generation unit. Block 18 is the entry point into the flowchart. In block 18 the program enters the process and descends into block 20 whereat a check is made to see if initialization needs to be performed. If initialization is to be performed, the process enters block 22 whereat the Slotted Token (SLT) with the format set forth in Figure 2 is generated. The process then descends into block 24 whereat the input port's ID in each of the sub-fields is set to an initial value. The program then descends into block 26 whereat the space reserved for writing available resources of an input port is initialized to 0. The process then enters block 28 whereat the SLT is forwarded and the program loops back to block 20. If in block 20 the initialization process was successfully completed the program descends into block 30 whereat it tests for arrival of the SLT. If the SLT has arrived the program then descends into block 28. If

the SLT has not arrived the process exits block 30 along the No path into block 20.

In this operation, the Token Generation Unit (TGU) generates the Slotted Token which is in constant circulation as long as the system is up. It is assumed that the ring provides a reliable transport mechanism so that the Slotted Token does not get lost or corrupted. As stated previously, each slot in the Slotted Token is associated with a single input port and indicates the available resources for the input port. Therefore, for InP:i the resource is shown as AvResInP:i in the Slotted Token. Initially when the TGU generates the SLT the TGU sets AvResInP:i to 0 for every InP:i. It is further assumed that the amount of available resources for an input port can be represented with a scale value greater than or equal to 0 or any other quantitative expression selected by the designer. When a node's Resource Control Unit receives the Slotted Token the Resource Control Unit updates AvResInP:i for every InP:i that it has. For example, assume, Node:1 (Figure 1) has allocated 150 (units of resource) for InP:1, 100 for InP:2, and 200 for InP:3. When the RCU for node 1 receives the SLT for the first time, it sets AvResInP:1 to 150, AvResInP:2 to 100 and AvResInP:3 to 200.

When a node's RCU receives the SLT it also uses that to reserve resources in other nodes to which it needs to send data for processing. For example, assume Node:0 (Figure 1) needs to send data, now or in the future, to InP:(K-2) or Node:(N-1). So, when the RCU for Node:0 receives the SLT it reserves resources in InP:(K-2) by deducting from AvResInP:2 the amount that it needs. For example, if the RCU for $Node_0$ wants to reserve 10 units in InP:(K-2) it deducts 10 from AvResInP:(K-2) before forwarding the SLT. It should be noted that the amount of resource reservation is limited to what is indicated as being available in the Slotted Token message.

Figure 4 shows a flowchart for the operation of the Resource Control Unit. The Resource Control Unit could be implemented as a state machine, a program processor combinatorial logic or similar devices. The flowchart in Figure 4 can be used to generate the Resource Control Unit as set forth in the specification.

Figure 4 shows a flowchart illustrating the operation of the resource control unit (RCU). The flowchart can be used by one skilled in the art to design the Resource Control Unit. In block 28 the program enters the process and descends into block 30 whereat a check is made to

see if the Slotted Token (SLT) has arrived. If the SLT has not arrived the process exits along the No path into block 52 whereat the process check to see if any data received at the input port (InP) has been processed. If the answer is No the process loops back into block 30. If the response is Yes the program enters block 50 whereat the resource released as a result of processing frames received at the port is added to the resource available for that particular port. The process then loops from block 50 to block 30.

Still referring to Figure 4, if at block 30 the answer is Yes the program enters block 32 where it checked to see if SLT has arrived for the first time. If the response from block 32 is No the process descends into block 36. In block 36 the total amount of released resource (ReResInP:i) associated with the input port InP:i since the last time the SLT was received is added to the amount of available resource for the input port InP:i.

From block 36 the program enters block 38. In block 38 the total amount of released resource (ReResInP:i) associated with the input port InP:i since the last time the SLT was received is reset to 0.

With respect to block 32 if the SLT is being received for the first time in the RCU the program enters block 34 whereat the RCU inserts the value for the resources available at each of its input ports and descends into block 38.

Still referring to Figure 4, from block 38 the program descends into block 40 whereat the RCU checks to see if a reserve resource needs to be cancelled. If the response is Yes the program exits block 40 along the Yes path into block 42 whereat for every port's reservation to be cancelled the amount is added to the resources available for that particular port. From block 42 the program descends into block 44. With respect to block 40 if the RCU does not desire to cancel a reserved resource the program descends into block 44. In block 44 the RCU decides if it needs to reserve a resource. If the answer is Yes the program descends into block 46 whereat the Resource Control Unit deducts from every input port the amount it needs from the resource available at the particular input port and enters into block 48. In block 48 the SLT is forwarded and the program loops back to block 30 to repeat the described process.

Therefore, a node cannot use another node's resource unless it performs an explicit resource reservation as described herein. A node should not use another node's resource beyond what it has reserved. When a node "consumes" its reservation it needs to make a new reservation if more resources are required. In subsequent receipt of the SLT (that is, after the first time) the RCU updates the AvResInP:i value in the SLT for every InP:i that it has according to the following scheme: Let ResInP:i be the total amount of release resource associated with InP:i since the last time the SLT was received. For every InP:i the RCU adds ResInP:i to AvResInP:i in the SLT before forwarding the SLT.

For example, if Node:1 releases five units of resource as a result of processing the data received at InP:(K-2), it adds five to ResInP:(K-2) which will be added to AvResInP:(K-2) the next time the RCU for Node:(N-1) receives the SLT.

A node may cancel its reservation for resources it has reserved in other nodes. When a node's RCU receives a SLT it also uses that to cancel any reserved resources in other nodes that it may not need. For example, assume $Node_0$ (Figure 1) has reserved ten units in InP:(K-2) and wants to cancel four units. When the RCU for Node:0 receives the SLT it adds 4 to the value of AvResInP:(K-2).

A decentralized "Advertisement-based" scheme is provided: Receivers advertise their resources and senders take what they need in a distributed fashion.

The invention supports dynamic resource reservation for

- One sender one receiver,
- One sender many receivers,
- Many receivers one sender, and
- Many senders many receivers

with one reservation message. Receivers need not know who the senders are when resource reservations are made. Can work inband or out-of-band.

CLAIMS

1. A method to reserve resources in a communications system comprising the acts of:

(a) receiving in a node a message having at least one sub-field to carry identifying indicia for at least one port and associated space for carrying information associated with resources available at said at least one port;

(b) examining the message with a resource control unit;

(c) if said sub-field carries an identifying indicia that matches an ID (identification number) of an input port in said node write in the associated space resources available at said at least one port, if the identifying indicia matches a port ID for a port to which the node wishes to communicate adjusting resource information recorded in the space to reflect resources reserved by said node.

2. The method of Claim 1 further including the act of transmitting the message.

3. The method of Claim 1 wherein the act of adjusting includes subtracting a scalar value from a value carried in said space.

4. A node for use in a communications network which carries out the method of any preceding claim.

5. A system comprising:

a plurality of nodes, including the node of claim 4; and

interconnect medium operatively interconnecting the plurality of nodes.

6. A computer program product comprising computer program code stored on a computer readable storage medium which, when executed on a data processing system instructs the data processing system to carry out the method of any preceding method claim.

FIG. 1

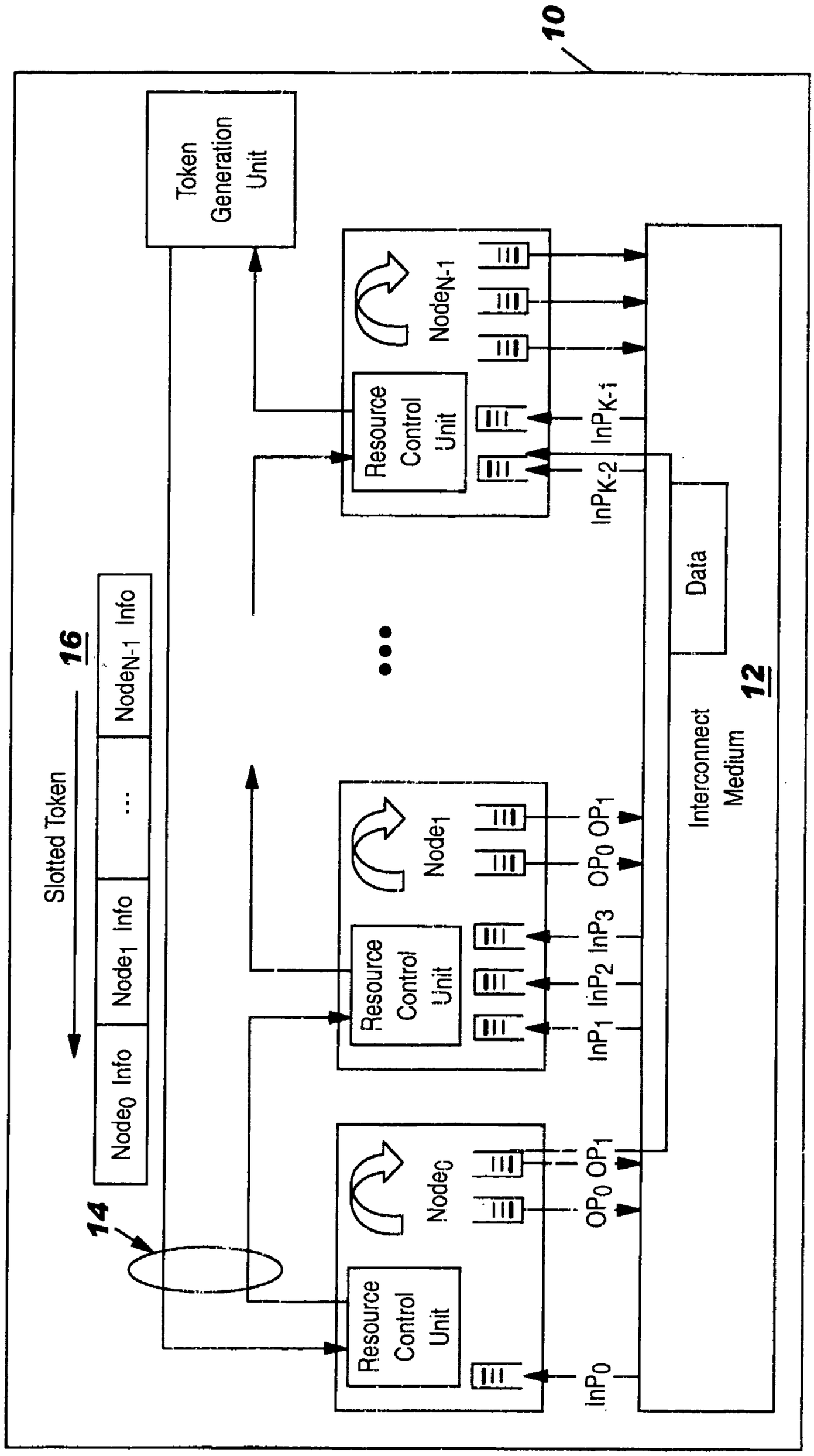

10
Token Generation Unit
Node$_{N-1}$
Resource Control Unit
InP$_{K-1}$
InP$_{K-2}$
Data
Slotted Token
16
Node$_0$ Info
Node$_1$ Info
...
Node$_{N-1}$ Info
14
Node$_1$
Resource Control Unit
OP$_0$ OP$_1$
InP$_1$ InP$_2$ InP$_3$
Node$_0$
Resource Control Unit
InP$_0$
Interconnect Medium
12

FIG. 2

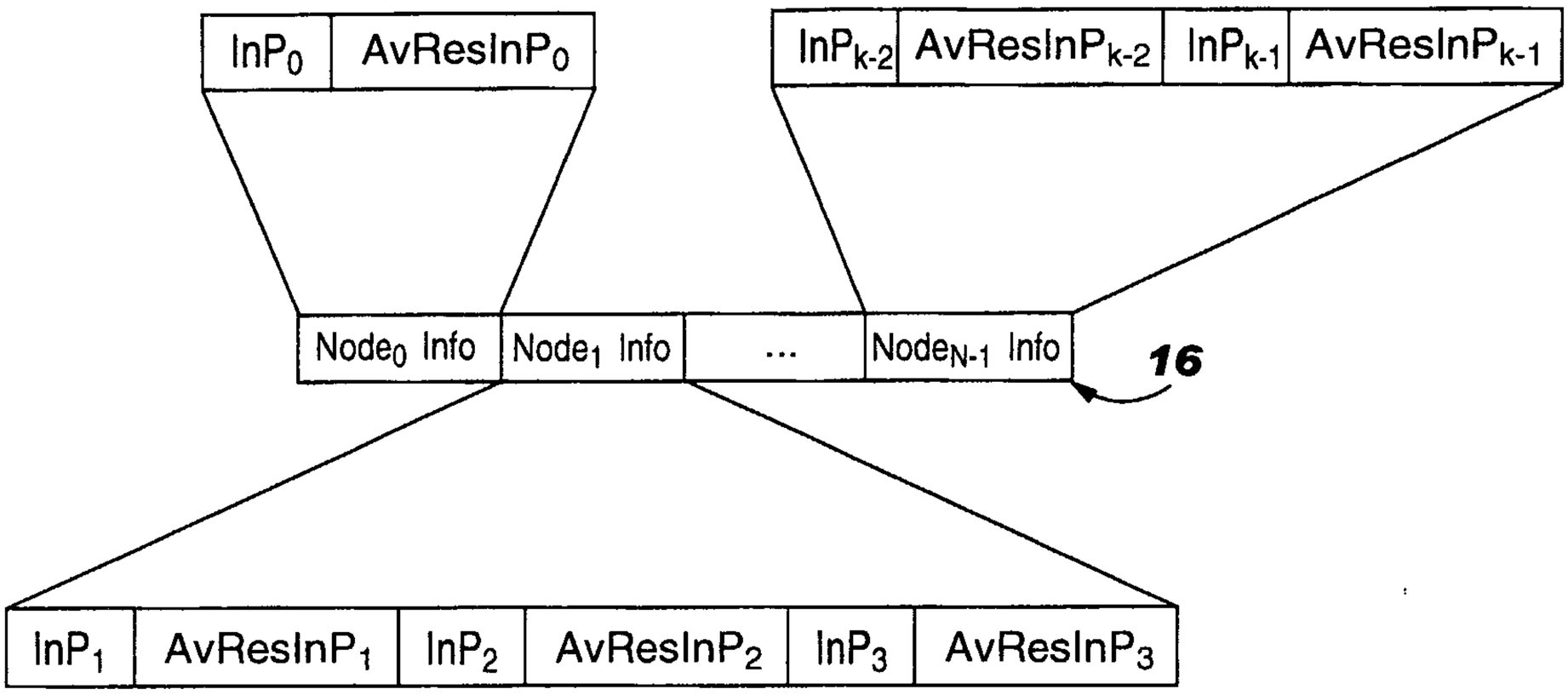

InP0
AvResInP0
InPk-2
AvResInPk-2
InPk-1
AvResInPk-1
Node0 Info
Node1 Info
...
NodeN-1 Info
16
InP1
AvResInP1
InP2
AvResInP2
InP3
AvResInP3

FIG. 3

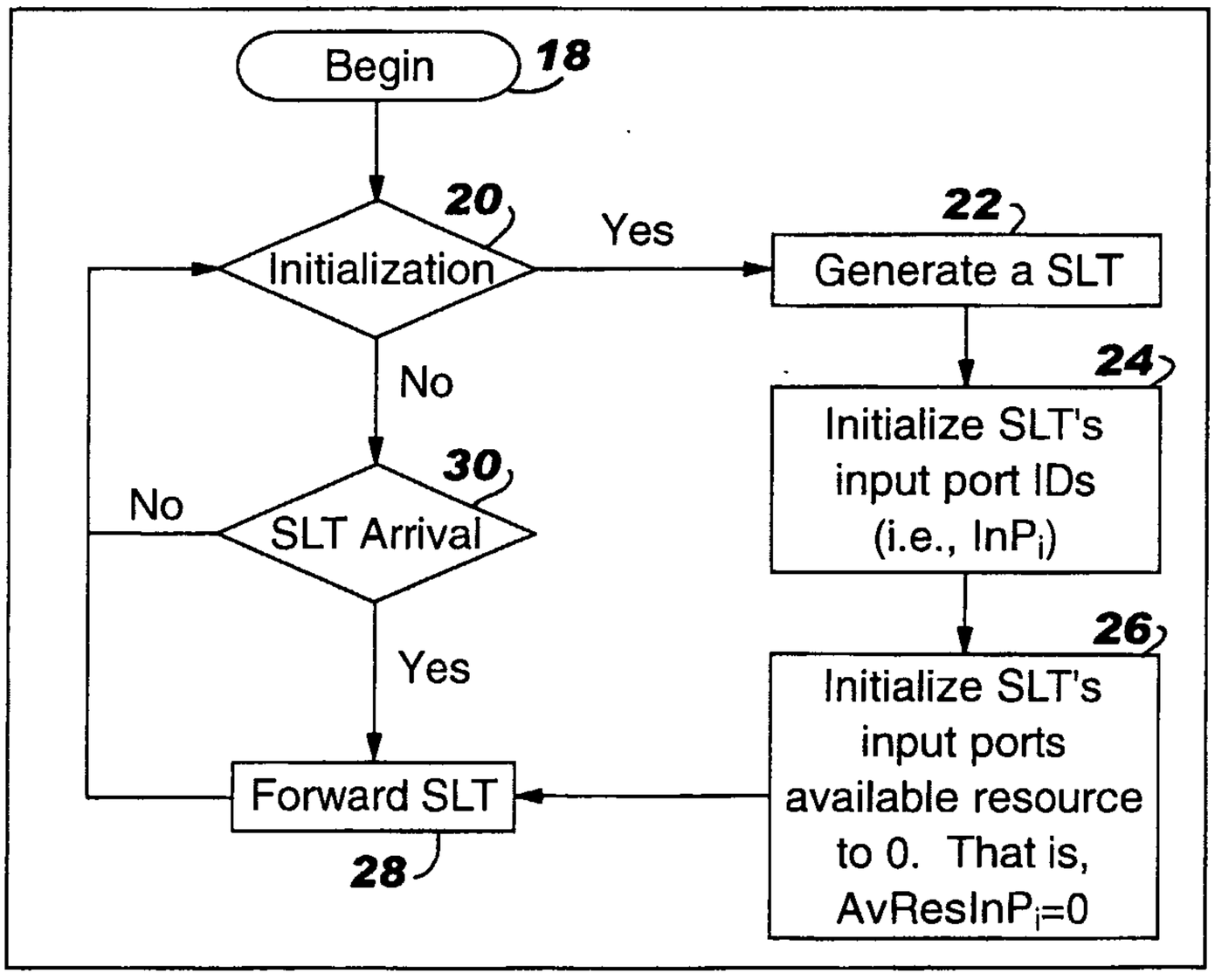

Begin
18
20
Initialization
Yes
22
Generate a SLT
24
Initialize SLT's input port IDs (i.e., InPi)
26
Initialize SLT's input ports available resource to 0. That is, AvResInPi=0
No
30
SLT Arrival
No
Yes
Forward SLT
28

FIG. 4

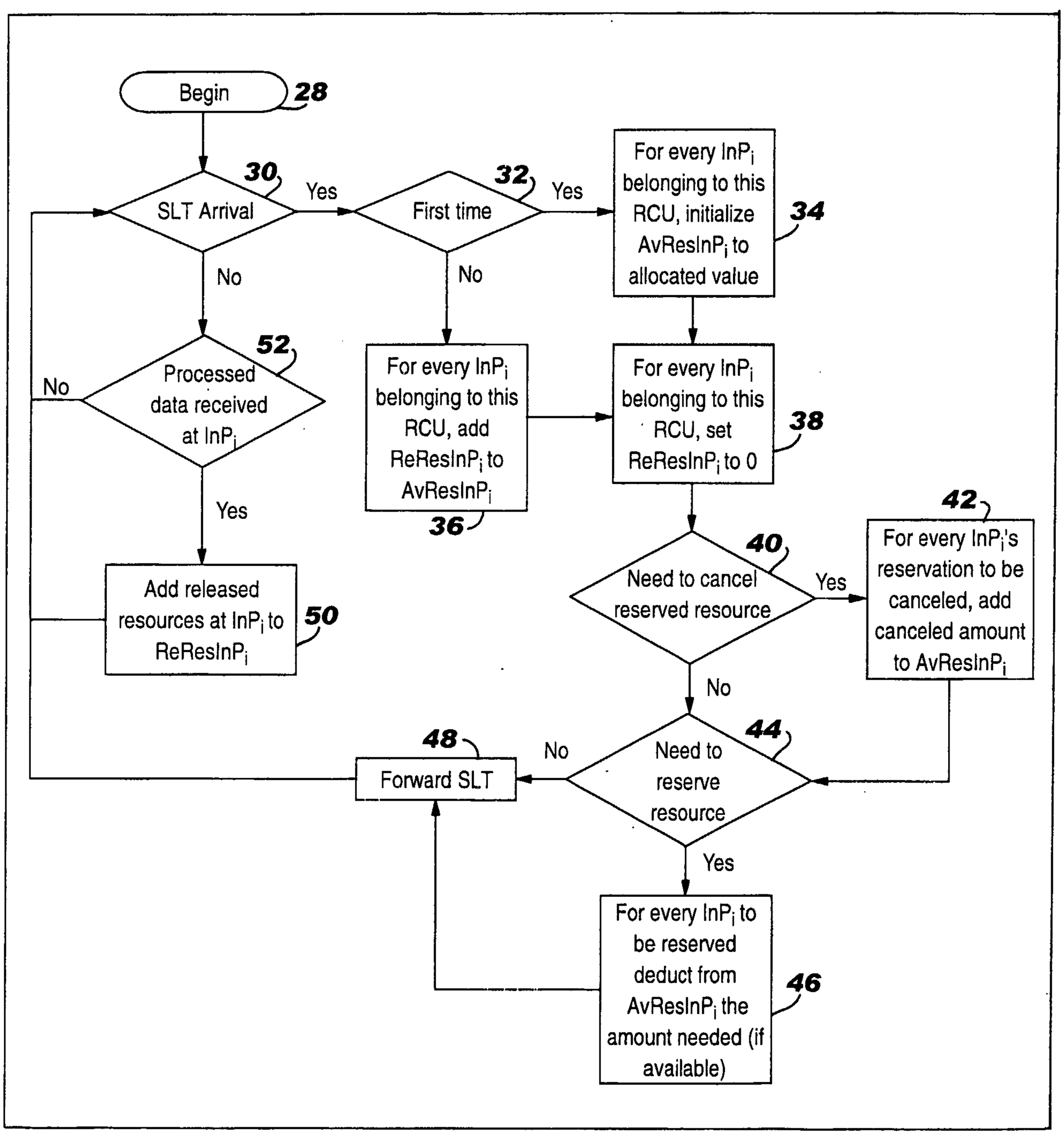
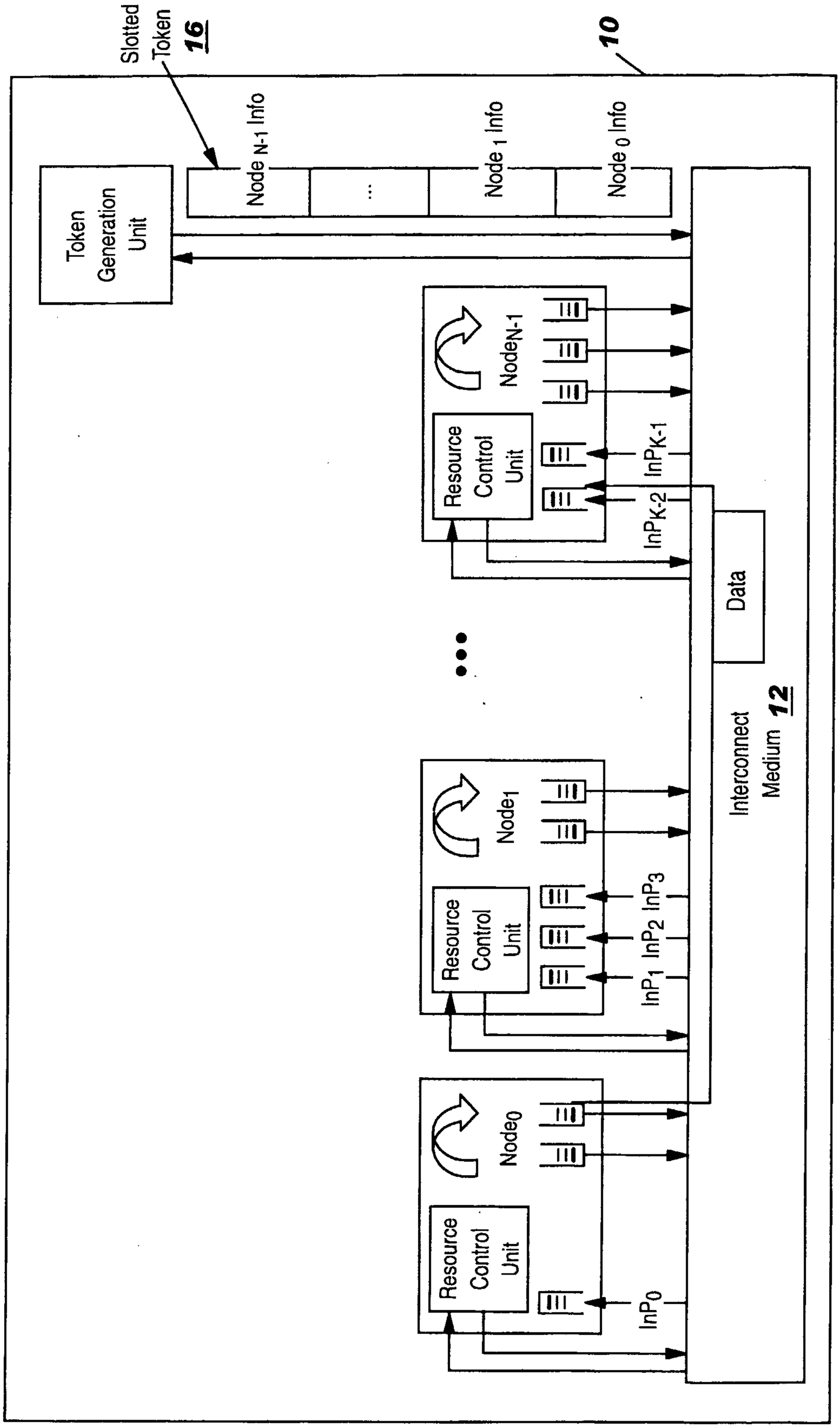
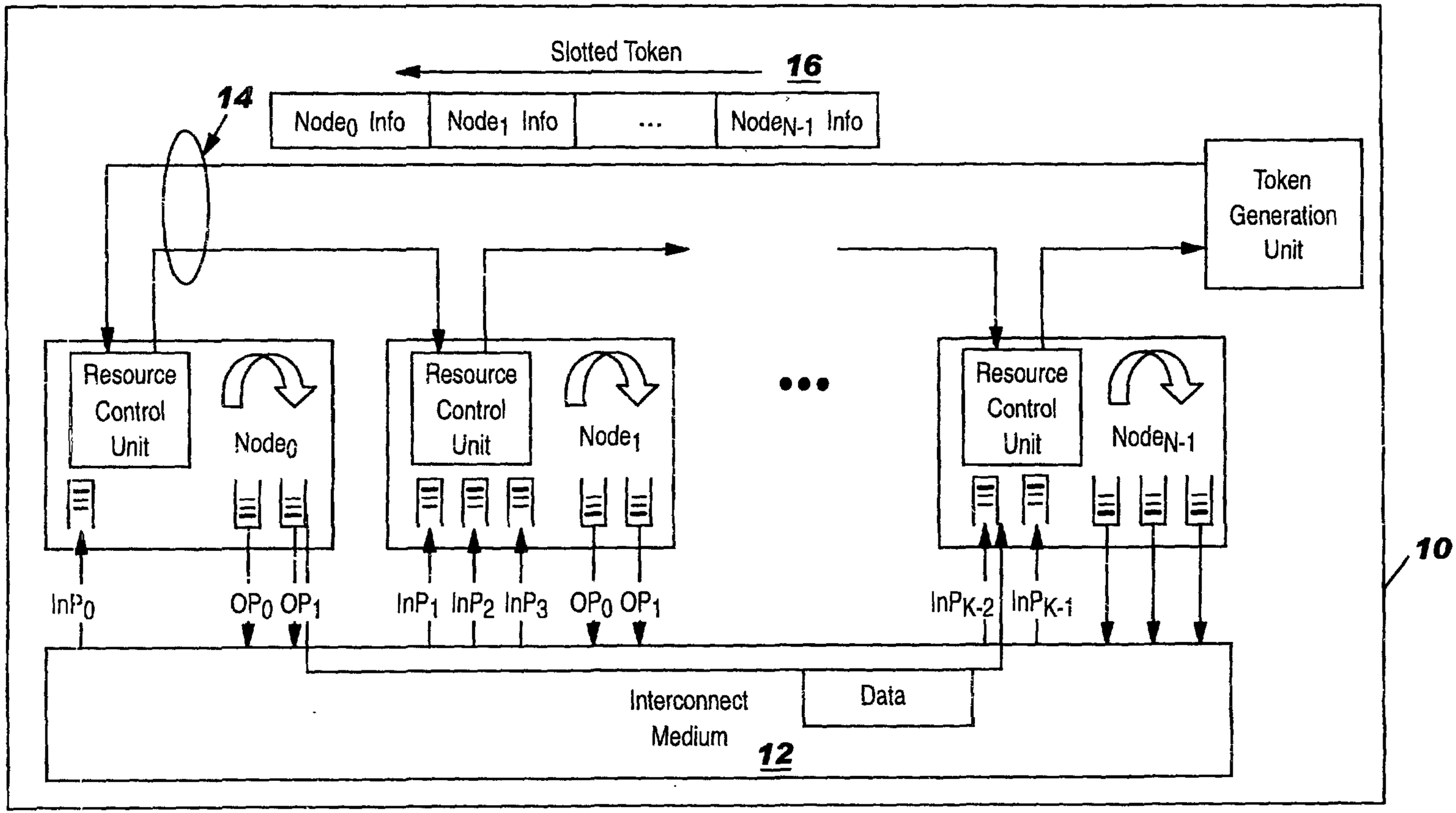

Begin
28
SLT Arrival
30
Yes
First time
32
Yes
For every InP_i belonging to this RCU, initialize $AvResInP_i$ to allocated value
34
No
No
Processed data received at InP_i
52
No
Yes
For every InP_i belonging to this RCU, add $ReResInP_i$ to $AvResInP_i$
36
For every InP_i belonging to this RCU, set $ReResInP_i$ to 0
38
Add released resources at InP_i to $ReResInP_i$
50
Need to cancel reserved resource
40
Yes
For every InP_i's reservation to be canceled, add canceled amount to $AvResInP_i$
42
No
Need to reserve resource
44
No
Forward SLT
48
Yes
For every InP_i to be reserved deduct from $AvResInP_i$ the amount needed (if available)
46

FIG. 5


Slotted Token 16
10
Node N-1 Info
Node 1 Info
Node 0 Info
Token Generation Unit
Node N-1
Resource Control Unit
InP K-1
InP K-2
Data
Interconnect Medium 12
Node 1
Resource Control Unit
InP 1 InP 2 InP 3
Node 0
Resource Control Unit
InP 0


Slotted Token
16
14
Node0 Info
Node1 Info
...
NodeN-1 Info
Token Generation Unit
Resource Control Unit
Node0
Resource Control Unit
Node1
Resource Control Unit
NodeN-1
InP0
OP0 OP1
InP1 InP2 InP3 OP0 OP1
InPK-2 InPK-1
Interconnect Medium
12
Data
10